DOCUMENT RESUME

ED 333 540 EA 023 031

AUTHOR Pantili, Linda; And Others
TITLE Principal Assessment: Effective or Not? A Meta-Analytic Model.
PUB DATE Apr 91
NOTE 43p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Administrator Effectiveness; *Administrator Evaluation; *Assessment Centers (Personnel); Effect Size; Elementary Secondary Education; *Evaluation Criteria; *Meta Analysis; Models; *Principals
IDENTIFIERS *National Association Secondary School Principals

ABSTRACT

The effectiveness of assessment centers in evaluating desirable criteria for the principalship is examined in this report, with a focus on the identification of the most effective evaluation criteria. This study explored a variety of factors in the preparation of school personnel for assessment, the actual National Association of Secondary School Principals (NASSP) assessment process, and the outcomes of the process. Methodology involved a series of meta-analyses to integrate factors internal and external to the NASSP assessment center process into a consistent theoretical framework for the improvement of principal preparation, assessment, selection, and performance. A model for future assessment center programs is proposed. Data were collected from 32 published and unpublished studies. Quantitative analysis computed the effect sizes of variables, and qualitative methodology generated an importance ranking of skill dimensions and a content analysis of common variable concepts. Findings indicate that leadership is the single greatest factor that affects administrative potential; race and gender had insignificant effects on performance dimension scores. In general, findings substantiate the validity of the assessment process in construct, content, and criteria. Recommendations are made for improving the quality of data for validity studies and for revising current preparation and assessment programs to more accurately reflect performance expectations in actual principal selection. Two figures are included. Appendices include a description of the studies, categorization of study types, and variables in the meta-analytic model. (38 references) (LMI)

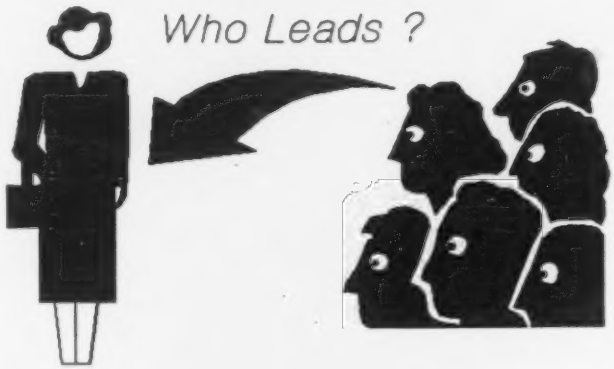
****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED333540

Principal Assessment: Effective or Not ?

A Meta-Analytic Model

Linda Pantili, John Williams, Jim Fortune
Virginia Polytechnic Institute
and State University



American Educational Research Association
Annual Meeting

Chicago, Illinois
April 4, 1991

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L. Pantili

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

EA023231

The Effectiveness of the Assessment Center Process:
A Meta-Analytic Model

Objectives

Are assessment centers effective in evaluating desirable criteria for the principalship? If so, which criteria are the most effective? If not, which criteria might augment the current assessment process? The National Association of Secondary School Principals Assessment Center Program has existed for nearly two decades and has assessed thousands of school personnel for selection and placement in principal positions. The assessment process has survived repeated challenges to validity and is enjoying increased service demand from states seeking to improve assessment and selection processes for school administrators. This study explored a variety of factors in the preparation of school personnel for assessment, the actual NASSP assessment process, and outcomes from the assessment process.

Theoretical Perspective

The latest validation of the NASSP assessment center process suggests this tool to be reliable, content valid, and valid when compared to external criteria for evaluating potential school leaders (Schmitt, et al, 1990). This follows a long series of evaluation studies examining the validity and efficiency of the assessment center process in recommending selection and placement of administrators in a variety of states: Arizona (Bley, 1983); Maryland (Farmer, 1985); Michigan (Blanck, 1989); Nebraska (Holman, 1986); New Jersey (Tryjankowski, 1990); Oklahoma (Ehinger, 1986); South Carolina (Crumpton, 1976); and Virginia (Van Newkirk, 1984).

However, these studies identified one or more portions of the assessment process as differentially affecting parts of the overall placement recommendation, as well as different position types of participants. Hence, multiple criteria were used in the related studies. As the use of the NASSP Assessment Center Program expands, the unexplained inconsistencies in program outcomes could increase reservations regarding process validity. A meta-analysis of the assessment process, *in toto* would more completely describe and identify the most desirable set of criterion factors associated with preparation and selection programs, staff development programs, placement assessment recommendations, and changes in administrator perspective after experiencing the assessment center procedure.

The NASSP Assessment Center program has been initiated in several states as a voluntary or mandatory procedure for the identification and selection of potential school administrators. Validity studies have followed in virtually every state in which the assessment process has been instituted. Concomitantly, additional independent research has identified different sets of factors important in preparing school personnel for placement in leadership positions (Thornton and Byham, 1982).

Factors including oral and written communication, stress tolerance, leadership, educational values, sensitivity and other psychological and skill variables have been associated with the performance of school principals and are measured in the assessment process. Other variables like participant attitude, administrator preparation programs, and self improvement efforts have also been identified in the literature but are not included in the NASSP center program.

Research has also indicated that principal preparation programs often fall short of teaching the skills identified as important to principal performance and assessed in the NASSP program (Engel, 1989). Coincidentally, the integration of assessment center performance factors into administrator education programs has been recommended to more closely link the preparation, assessment, and performance of principals (Wendel, 1988).

Yet, post-assessment surveys of administrator attitudes (Ford, 1987) and self improvement efforts (Walden, 1985) suggest that participants perceive the factors of performance to be significantly different from what they learned in preparation programs or were assessed for in the NASSP program. These discrepancies between preparation, assessment and actual job performance may have enticed many employers of principals to rely more heavily on traditional selection procedures--interviews and in-house training (Pokorny, 1985).

In summary, research studies to date have been individually inconclusive in identifying the best criteria for assessment and selection of school administrators. This study used a series of meta-analyses to integrate the factors identified within and outside of the NASSP assessment center process into a consistent theoretical framework for the improvement of principal preparation, assessment, selection and performance. A model for future assessment center programs is proposed on the basis of this analysis.

Methods of Study

Selection of Studies

A comprehensive literature review utilizing the ERIC data base, Dissertation Abstracts, Intl., NASSP studies and professional references was conducted to collect studies related to principal preparation, assessment and performance connected to the NASSP Assessment Center Program. No studies were excluded due to quality considerations and no rating of study characteristics was attempted beyond simple categorization. We found that all studies utilized non-experimental designs and more than one design in each study, usually in an attempt to describe as many of the characteristics of the NASSP assessment center process as possible. As a result, inconsistency is a primary characteristic of studies used in this analysis.

Data

Data for meta-analysis came from 32 published and unpublished studies of assessment center validity, principal performance comparison, and administrative attitude evaluation. Correlations, means, standard deviations, and rank order data were drawn from published and unpublished sources on all variables believed to affect principal preparation, assessment, and performance. These studies and variables are summarized in Appendices One, Two and Three.

Most studies of assessment center processes and factors associated with principal preparation programs provided different components of the study data base. Examples included were in the form of correlations among rating scales, correlations between recommendations and external criteria, t-tests and ANOVA on race, gender and personnel classification averages, and ordinal comparisons of attitude measures. Variables in the data bases which could be compared across studies were used to integrate effect sizes from separate evaluations.

Analytic Procedures

Assessment center variables selected for meta-analysis were skill ratings, gender, race, position type and principal performance measures. Variables selected from outside the assessment center process were attitude, self-improvement, and superintendent ratings of administrative applicants. These variables provided the basis for four separate quantitative meta-analyses of gender, race, and position type effects on NASSP assessment center skill dimensions and validity effects between NASSP dimensions and external performance criteria. Additionally, two qualitative syntheses were generated with importance ranking data of skill dimensions from various studies and a content analysis of study findings.

I. Quantitative

All the quantitative data were converted to unbiased estimates of standardized "r" effect size scores using methods reviewed in Wolf (1986) and Rosenthal (1984) (See Table One for formulae used). These were converted to standardized effect sizes (Z_r) before being combined into average effect sizes from various studies. "Effects" for meta-analysis variables were considered independent across studies and assumed so when criterion validities were combined across multiple criteria within single studies. The homogeneity of effect sizes was tested to determine the appropriateness of combining effect sizes from studies being analyzed.

TABLE 1
Meta Analysis Formulae Used for Effect Size Estimation

Original Statistic	Conversion to "r"
t	SQRT((t^2)/(t^2 + df))
F	SQRT((F)/(F + df error)) where F(1,df error)
Pearson r	r
Conversion to Z_r:	1/2LOG(base e)[(1+r)/(1-r)]

Effect sizes were combined across studies within gender, race, and position type variables and across study findings on performance criteria variables by first generating Fisher transformations on "r" effect sizes and then computing weighted (by df) and unweighted mean effect sizes. Unweighted effect size variance estimates were also computed. Reliabilities of assessment center measurement variables were not used in this study to adjust effect size estimates, due to the lack of enough study data to provide stable reliability estimates (we sought reliability estimates beyond the few NASSP estimates discovered).

II. Qualitative

A qualitative synthesis was performed by comparing importance rankings of NASSP skill dimensions done by the authors of twenty-six studies. Additional ranks were created for other studies by ranking reliability reports and criterion validity coefficients as "importance" criteria. A summation of ranks for each dimension was computed and used to weight each dimension over all other dimensions in a "master" ranked list ranging from 13 (most important) to 1 (least important), (See Figure One in Results).

A second qualitative synthesis was generated from a content analysis of all study findings for common concepts on variables investigated. The askSam™ (AskSam Systems, Inc., 1989) software system was used to create a data base of study findings and perform the content analysis, based upon "hypertext" linkage of common variables. In addition, Lights' (1984) strategies for clustering literature concepts were used to refine the summary of study findings.

Large Sample studies

Our sample size distribution was positively skewed due to the presence of one large national study (Schmitt, 1990) with a sample of some 2800 individuals. Mean effect sizes were computed with and without the data from this study, but no substantive differences resulted. Final results are presented which include the national study sample.

Results

Quantitative analyses precede qualitative results in this study. Descriptive statistics and categories for the studies used in this meta-analysis are presented in Appendices. Appendix One provides a summary of the characteristics of each study. Appendix Two categorizes studies into descriptive and inferential types broken down according to precursor, process and criterion variable definitions. Appendix Three presents the variable list used in the analytical procedures.

The first and most important meta-analysis averaged the effects from 47 findings in 10 different studies of various measures of principal job performance used to validate the NASSP assessment center process. The criteria used in the studies to validate NASSP skill dimensions were job performance criteria ranging from conceptual levels to various administrative functions, such as fiscal management and other employment-related issues (See Appendix One). Combined effect sizes of criterion validity on the 12 skill dimensions and overall assessment center rating are presented in Table One below.

Surprisingly little association (average effect r<= .20) exists between the NASSP assessment ratings and job performance criteria for principals. Several of the skill dimensions used in assessment appear to have near zero validity with job performance criteria (e.g., sensitivity, range of interests and personal motivation). Perhaps, these skill dimensions could be modified to better match principal job performance criteria.

Skill dimensions seeming to have more effect on job performance criteria are organizational ability, leadership and communication skills. However, none of the dimensions, nor the overall ratings, seem to share more than a few per cent of total variance with job performance criteria used in assessment validity studies. Either better criteria for principal job performance need to be developed or NASSP assessment center skill dimensions should be changed to reflect a higher association between assessment center measurements and validating criteria.

TABLE 1
Combined Effect Sizes Across Job Performance Criteria Validities

Unweighted K=47	Unweighted Mean Z_r	Unweighted variance Z_r	Weighted Mean Z_r	r
Problem Analysis	.197	.092	.179	.198
Judgment	.157	.055	.108	.158
Org. Ability	.251	.034	.247	.257
Decisiveness	.166	.052	.137	.168
Leadership	.251	.039	.246	.257
Sensitivity	.091	.051	.012	.091
Range of Interests	.089	.041	-.002	.060
Personal Motivation	.066	.033	.050	.080
Education Values	.135	.012	.013	.070
Stress Tolerance	.216	.044	.105	.135
Oral Communication	.170	.114	.166	.221
Written Comm.	.239	.007	.109	.261
Overall Rating	.305	.139	.215	N/A

N/A: not applied due to the lack of sufficient numbers of studies (k=2).

Effects were also combined across studies for gender. The effect sizes on all twelve NASSP skill dimensions for male and female comparisons (t-tests) were combined from seven studies. Average effect sizes and variances are presented in Table Two.

Overall, the effect size estimates combined across studies suggest that gender does not play a significant role in assessment center scores on any dimension. Individual studies suggest significant statistical differences between female and male scores on some of the skill dimensions, comparison of substantive differences (which must equal one full scale point to change rating on the overall 1-5 scale) in those studies corroborates the meta-analysis findings of small or trivial gender effects.

TABLE 2
Combined Effect Sizes Across Gender

Unweighted K=7	Unweighted Mean Z_r	Unweighted variance Z_r	Weighted Mean Z_r	r
Problem Analysis	.207	.066	.214	.204
Judgment	.235	.076	.187	.232
Org. Ability	.210	.059	.207	.207
Decisiveness	.076	.011	.091	.076
Leadership	.176	.079	.143	.174
Sensitivity	.166	.072	.145	.164
Range of Interests	.066	.002	.084	.066
Personal Motivation	.194	.063	.196	.193
Education Values	.168	.052	.193	.191
Stress Tolerance	.127	.052	.116	.126
Oral Communication	.231	.078	.220	.228
Written Comm.	.288	.057	.376	.280
Overall Rating	.092	.009	.139	.092

Effect sizes were next combined (t-tests and ANOVA collapsed to two categories) across four studies for the race variable. This variable was defined as the categories white and nonwhite for purposes of the meta-analysis. Table Three summarizes the combined results.

Race seems to have a moderate ($r > .30$) effect size on performance dimension scores overall, especially on organizational ability and written communication skills. Again, this result is borne out by several of the individual studies in which nonwhite assessment participants scored significantly lower than white counterparts. However, differences do not appear to be substantively meaningful and this may account for the modest effect sizes noted above.

TABLE 3
Combined Effect Sizes Across Race

Unweighted K=4	Unweighted Mean Z_r	Unweighted variance Z_r	Weighted Mean Z_r	r
Problem Analysis	.312	.013	.363	.263
Judgment	.310	.023	.347	.259
Org. Ability	.410	.018	.395	.338
Decisiveness	.304	.027	.249	.269
Leadership	.296	.015	.289	.252
Sensitivity	.084	.001	.111	.077
Range of Interests	.148	.003	.157	.135
Personal Motivation	.169	.003	.099	.164
Education Values	.085	.007	.036	.075
Stress Tolerance	.159	.002	.194	.150
Oral Communication	.190	.004	.249	.189
Written Comm.	.394	.008	.275	.340
Overall Rating	.293	.022	.262	.285

Finally, effect sizes were combined across five studies for position types. Position types were defined as participants/non-participants, teachers/non-teachers and assistant principals/non-assistant principals. Combined effects are listed in Table Four. These effect size estimates must be considered less reliable than the preceding variable estimates since the combined effect sizes failed the test for homogeneity. Nonetheless, there does not appear to be much of an effect of position type on the skill dimension scores and certainly there is no substantive effect of any position type beyond one point on the five point rating scale used by the NASSP.

Rank Importance of Dimensions

The relative importance of each skill dimension in the NASSP assessment center process was analyzed by aggregating importance rankings across all studies (26) containing rankings based upon relationships to validity criteria (See Figure 1). These studies agree that Leadership is the single most important dimension to be measured in the assessment of administrative potential, accounting for over 20% of all importance rankings and being ranked first among all dimensions. This reflects the relatively consistent and moderate effect size across various performance criteria on the leadership dimension in the meta-analyses above.

TABLE 4
Combined Effect Sizes Across Position Type

Unweighted K=5	Unweighted Mean Z_r	Unweighted variance Z_r	Weighted Mean Z_r	r
Problem Analysis	.171	.021	.128	.173
Judgment	.195	.031	.117	.198
Org. Ability	.172	.010	.074	.174
Decisiveness	.135	.003	.129	.136
Leadership	.138	.001	.083	.139
Sensitivity	.092	.002	.066	.092
Range of Interests	.102	.003	.122	.102
Personal Motivation	.120	.002	.111	.121
Education Values	.123	.008	.099	.124
Stress Tolerance	.107	.001	.096	.107
Oral Communication	.121	.003	.089	.122
Written Comm.	.124	.009	.137	.125
Overall Rating	.156	.002	.051	.157

After the leadership dimension was ranked across all studies, all other dimensions tended to be subsumed under leadership, regardless of the criterion used for ranking. For example, organizational ability and judgment were the next most important dimensions cited in studies and both may be considered a part of the leadership construct. Oral and written communication ranked next in importance across studies and may also be subsumed under leadership skills. These five dimensions account for over sixty percent of all importance ranks assigned to dimensions by these studies. These same dimensions also account for the most consistent and highest effect sizes derived in the meta-analyses above.

Leadership, then, stands out as the single greatest factor affecting administrative potential, according to importance ranking and meta-analysis of study findings. Other important dimensions may be considered a part of the leadership skill dimension. This observation may support the results of several factor analyses of assessment center data, in which one, two or four factors were identified (Ogawa, 1986; Schmitt, 1990). We suspect that these factor analyses were identifying leadership and its associated dimension skills.

Content Analysis of Study Findings

A content analysis of study result reports was conducted to establish commonalities in the variable relationship conclusions drawn by various study authors, following thematic synthesis procedures suggested by Light and Pillemer (1984). The results are summarized around each variable topic below.

Are we using assessment centers for selection?

Studies reviewed indicate that school districts are not making wide use of the assessment center approach for principal selection. Instead, most school districts appear to be relying on the old standby, the interview process. Research findings from assessment center studies are generally not being implemented. Quotations from individual study findings follow:

> **"Most school districts see a need for training potential candidates for administrative positions... (Pokorny, 1985). Assessment centers are established for their expediency in providing a known pool of talent from which to select principals (Van Newkirk, 1984). One study reported that school districts rely heavily on interviews for selection purposes and not on techniques recommended in the literature (Pokorny, 1985). There has been very little change in the selection techniques used over the last five years (Pokorny, 1985)".**

What selection criteria have been addressed?

Few studies gave a specific list of the factors to be considered in the selection process for school administrators. Studies addressed the need for some type of selection criteria based on assessment center results, yet also recognized the limited application of any particular type of selection criteria from this process. For example:

> **"Selection criteria perceived by typical middle school principals were described as 30 to 44 years old, male, white, MS degree, teaching experience, leadership in curriculum development and curriculum change. Masters degrees were important for principal candidates; doctorate degrees are not necessary (Scholl, 1980). Principals and superintendents, regardless of district size and wealth, gender, and race were very similar in their perceptions with respect to selection criteria and the importance of assessment center factors" (Crumpton, 1989).**

Other descriptive information gleaned from the studies addressed the following with regard to principal candidates.

> **"Non-white candidates do less well than white candidates (Schmitt, 1981). "Outsiders" hired as principals had fewer difficulties in human relations than those familiar with the district, and principals are unlikely to function as change agents in the bureaucracy of school administration" (Alvy, 1983).**

Can preparation and training programs improve performance?

Review of the studies involved with preparation and training programs reveal a lack of concentrated efforts to develop the generic skills identified by NASSP. Internships appear to be the "best" method of instruction for school administrators and appear to increase the level of development of the twelve generic NASSP skills. For example:

> **"Principal preparation and training programs are largely conducted at the understanding level, with most of the generic skills developed to a only a moderate extent (Engel, 1989; Voit, 1989). One study indicated significant relationships between educational administration (ED AD) grades, peer rating, instructor rating, and the generic skills of problem analysis, judgment, leadership, and organizational ability. Indicators of success in job performance by school principals were problem analysis, promotion and leadership, organizational ability, human relations skills and personal motivation" (Jacobson, 1986).**

A study in Michigan (Voit, 1989) described training skills deemed necessary by principals. The "people factor" was credited as the most successful approach to principal training programs and the "teaching" of the generic skills. Studies suggested that:

> **"Principal respondents specifically asked for training and development opportunities for improvement in coping with and managing stress on the job, use of appropriate judgment, how to solve problems, and leadership. The "people" factors which contribute to successful training programs were identified as peer interaction, instructors with practical experience, and internships. Principals said the internship was the most ideal mode for developing 78% of the generic skills needed in their jobs" (Voit, 1989).**

Wendel's (1984) research report comparing universities who are members and non-members of the University Council for Education Administration provided little information on differences in preparation programs. Neither type of program

seemed to be donating energy and resources to advancing the use of assessment center information for skills development.

Wendel's report found differences in the use of the assessment center method in principal preparation programs between University Council for Educational Administration (UCEA) and non-UCEA member institutions at the departmental level to be negligible. For example:

> **"UCEA faculty members do report incorporating more of the assessment center dimensions integrated into their course work than the non-membered universities. Interest in establishing an assessment center expressed by department chairpersons and faculty at the universities was low".**

Who Performs Best?

Females often score higher on some particular skills and overall ratings compared to males. Yet, these higher scores (tenths of a scale increment) did not translate into substantive differences on the one to five rating scale used in assessment. Examples are:

> **"Females performed slightly better than males, reflecting job-related skill differences, not bias in the assessment center process (Holman, 1986; Schmitt, 1990). Female candidates scored higher than male candidates in all twelve of the skill areas and the overall rating in the assessment center, scoring significantly higher on sensitivity, organization and planning, oral communication, and written communication" (Dickson, 1987). On assessment activities, females scored significantly higher on in-basket and task direction and the same as males on the leaderless group discussion (Friedman, 1980). Females perform better than males utilizing supervisor and self ratings (Schmitt, 1990). Participants with supportive supervisors, especially females, were more likely to have higher overall performance ratings, perceive increased career opportunities, change jobs, and use assessment center results for professional development" (Dickson, 1987).**

Do higher conceptual level principals perform more effectively?

Holman (1986) and Micek (1983) performed comparative studies relating principal performance with conceptual levels. Their research did not support the concept of higher conceptual levels indicating highly effective principals. For example:

> **"Effect interaction was determined between gender and conceptual level and was specific to the skill dimensions of organizational ability, decisiveness, and written communication. High conceptual level females performed significantly better than high conceptual level males on organizational ability and written communication and better than low conceptual males on organizational ability" (Holman, 1986). No significant differences between performance for high and low conceptual levels of males was found for the skill dimensions. A significant interaction between conceptual level and occupational level was also identified, but no simple interaction effects were significant (Holman, 1986). One study reported no support for a prediction that principals with higher degrees of conceptual level were perceived by teachers to perform more effectively than principals with lower conceptual levels " (Micek, 1983).**

Do assistant principals have an advantage in the assessment center process?

The studies reviewed in this analysis indicate assistant principals tend to score higher than non-assistant principals in the assessment center process but not substantively higher on the overall assessment scale of one to five. Higher scores were attributed to administrative experiences that accompany this position. Studies suggested that:

> **"High career aspirations, career opportunities, and usefulness for professional development were linked to higher overall performance ratings (Dickson, 1987). There was a significant difference among the mean scores for five role groups of assistant principals, educational specialists/counselors, and teachers from the elementary, middle, and high school levels for five of the skills identified and the overall rating" (Blanck, 1989). Assistant principals perform better than non-assistant principals in the assessment center probably because of administrative type experiences. Candidates serving in non-teaching roles as counselors or educational specialists performed better than teachers in one study (Schmitt, 1981, 1990).**

How valid is the assessment center process?

Construct, content, and criterion validities were addressed in several studies. Generally, analyses of information in studies substantiates the validity of this process in all three categories. For example:

> **"Analysis indicates high interrater agreement with skill levels supporting internal validity and a meaningful difference among the various skills as well as agreement about candidates' skill levels. Assessors view the dimensions as measuring somewhat different skills but those performing high on one dimension perform well on other dimensions" (Schmitt, 1981, 1990). There was a significant relationship between assessment center skill ratings and job performance rating, with higher correlations between job performance ratings by principals' superiors compared to teacher and support staff performance ratings. Supervisor and teacher ratings were considered the most valid indicators of principal's job performance (Schmitt, 1981, 1990). There was no relationship between the assessment center scores of the principals and the teacher rating on the Administrative Competency Rating Scale, (ARCS)" (Britt, 1976).**

Will attitudes affect assessment center usefulness?

Ford (1987) and Bley (1983) studied participant attitudes toward assessment center results. High and low over-all attitude rating did affect use of results by participants. Ford found that:

> **"Attitude scores relative to low and high placement recommendations were significantly different in participants and non-participants toward the use of assessment centers as a reliable means of diagnosing strengths and weaknesses, related to gender in the dimensions of problem solving, decisiveness, and leadership; and were significantly different in the use of assessment centers as a reliable means of diagnosing strengths and weaknesses in the dimensions of judgment, decisiveness, leadership, range of interest, personal motivation and staff development" (Bley, 1983; Ford, 1987). Yet, no difference was found in attitude to identify potential candidates for administrative positions as related to gender. This study also reported no differences in the attitudes of participants and non-participants toward the use of assessment center as related to job classification for identifying potential candidates for administrative positions" (Bley, 1983).**

Farmer's (1985) study compared three categories of participants; those recommended for promotion, those recommended and not promoted, and those not recommended. Farmer expanded the type of information provided by Ford to trace opinions after the experience of the assessment center process in relation to what happened with regard to job placement of participants. Farmer discovered that:

"Three categories of participants disagreed on using the assessment center scores for promotion. Participants who were recommended, whether promoted or not promoted, agreed on the importance of an opportunity to observe modeling of administrative skills in their present jobs. Those participants who were not recommended disagreed significantly that they had the opportunity to observe the skill dimensions on the job and had the greatest difference of opinions regarding the influence of participation in the assessment center on positive feelings about the school system.

Significance of the Findings

Meta analysis of the assessment center validity studies encountered to date suggests that significant improvements in the quality of data available for validity studies is called for. Most of the studies done thus far have been descriptive and qualitative in design (See Appendix One). Researchers seem to be using unique criterion variables every time a new study is commissioned. A uniform criteria needs to be established for the validation of relationships between assessment center ratings and administrative job performance.

This analysis also illustrates the need to revise current preparation and assessment programs to more accurately reflect performance expectations in actual principal selection. Moreover, it appears that less relevant variables like gender, race, and position type may be dropped as precursor aspects in the consideration of individual results. Results of criterion validity studies on job performance are somewhat ambiguous and effect sizes of the NASSP skill dimensions are relatively low across studies.

Synthesis of findings and rankings suggest that skill dimensions are highly inter-related and probably measuring one or, at most, two primary behaviors. This "factor" may be identified as leadership. If so, it is not surprising that criterion effect sizes are relatively small since the effects of one skill dimension may be confounding the effect size of another. Relationships between various performance criteria may likewise be of major importance in reducing effect sizes.

Suggestions for Improvement

1) Several of the current skill dimensions may be pulled out of consensus ratings and treated as descriptive information: e. g. range of interest, personal motivation, educational values, and stress tolerance. These are skill dimensions with the lowest effect sizes across the studies.

2) Factor analyze the current skill dimensions to identify components that would contribute to a more effective and accurate assessment center model (See Figure 2). Figure two suggests that the twelve skill dimensions already identified and other "unknown" causal factors are acting through the performance categories of **curriculum and instruction, personnel, and administration** to effect the performance required of potential administrators. An improved model would focus on these three areas identified from the meta-analysis of the ranked performance criterion. This model would seem to better address the administrative performance criteria and might more appropriately identify candidates for administrative preparation programs and professional development programs.

3) This meta-analysis reveals that demographic variables are not strong effect agents. Future studies might ignore background factors such as gender, race and position type.

4) Relatively small effect sizes associated with the NASSP skill dimensions and current performance criteria suggest that a study of performance criteria be conducted and a data base be established to determine and utilize those performance criteria which validate the NASSP assessment process.

These recommendations for improvement of the NASSP Assessment Center Process will hopefully lead to the further application of meta-analyses to the refinement of assessment models. In addition, the recommendations may lead to a better measurement model of administrative potential and a better validated assessment process. With an improved assessment process, we anticipate NASSP Assessment Centers will reach a higher potential to serve the purpose of identifying potential administrators, refining principal preparation programs and enhancing professional development.

DIMENSION MODEL
Percent of Importance Across Studies

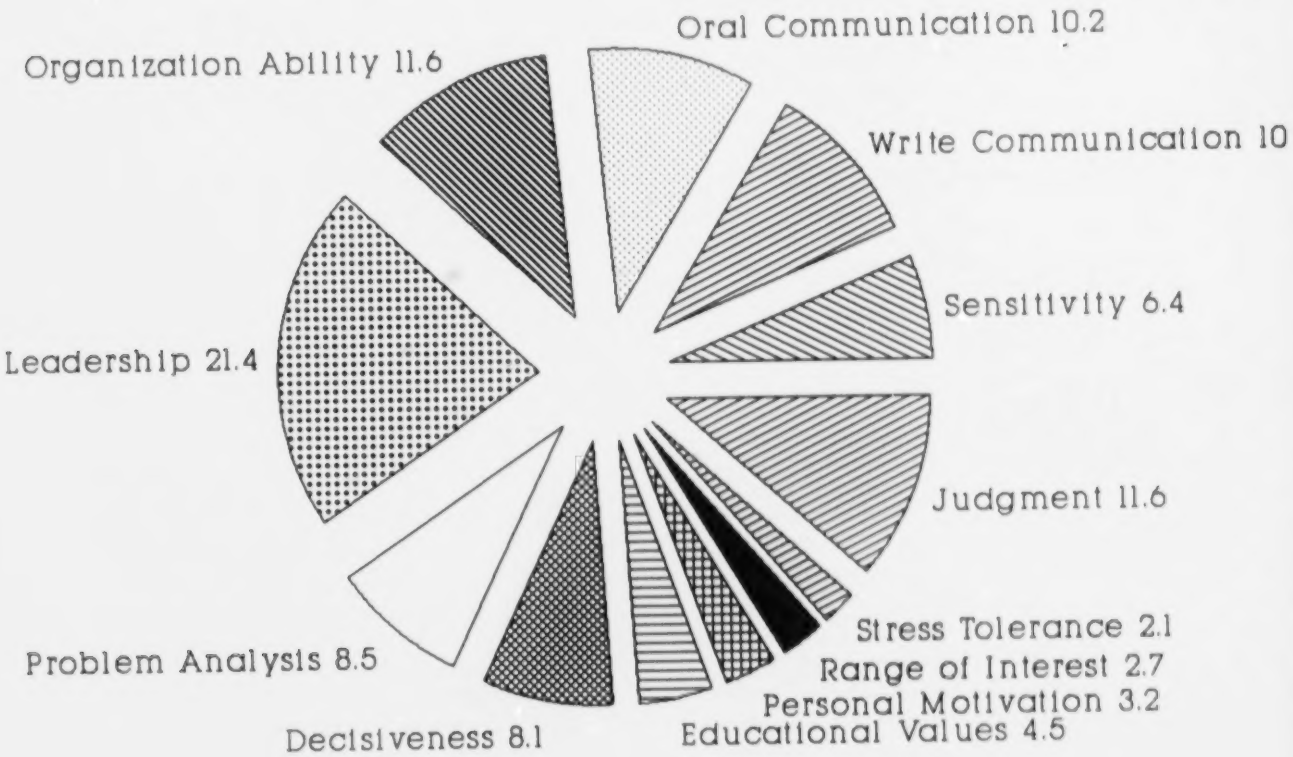


Figure 1

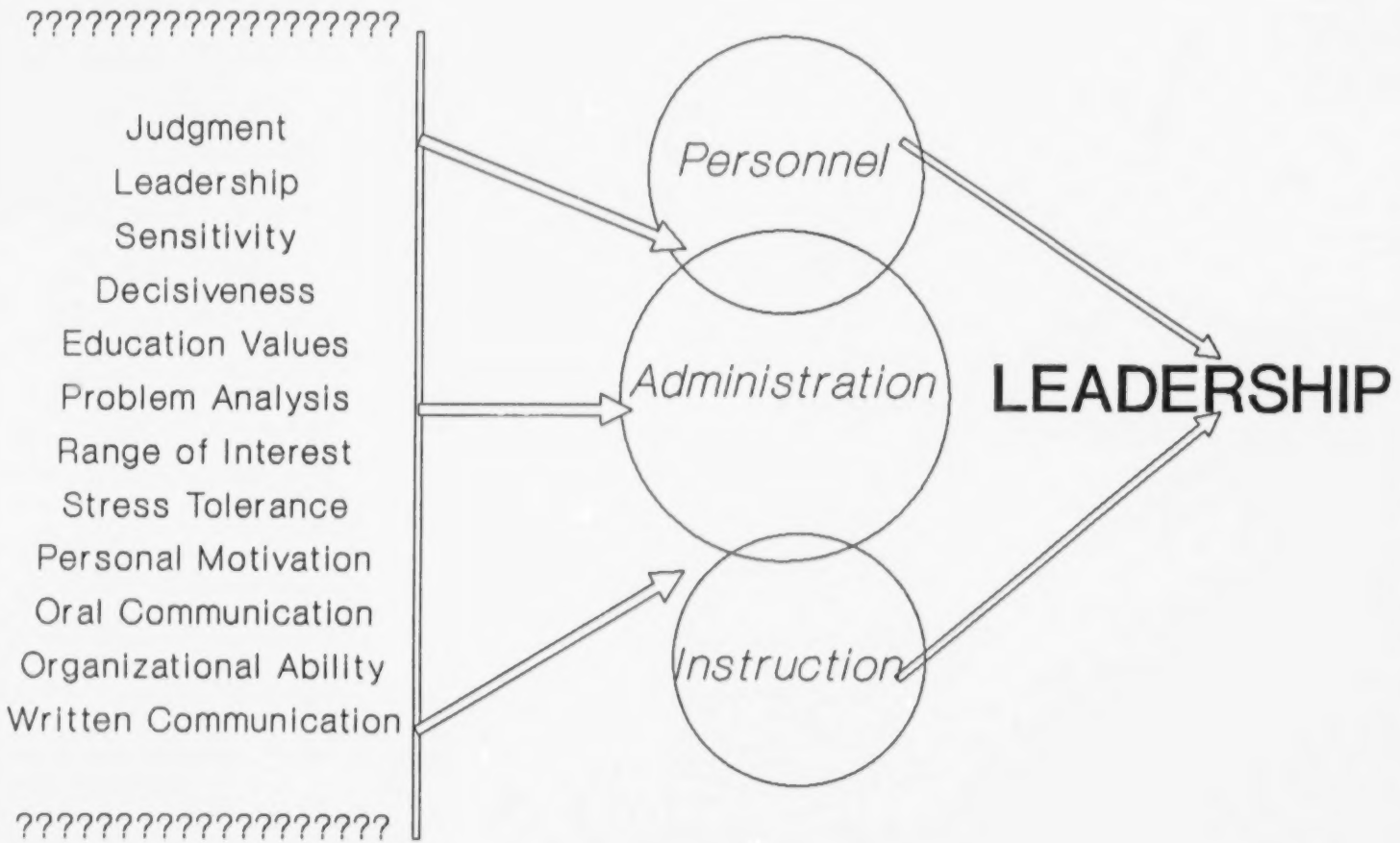


Figure 2. **A Revised Model for Principal Assessment**

References

Alvy, H. B. (1983). The problems of new principals. Doctoral dissertation, University of Montana.

AskSam (1989). AskSam software system. Perry, Florida: AskSam, Inc.

Blanck, J. I., (1989). Comparison of five role groups of michigan participants in nassp assessment centers. Doctoral dissertation, Western Michigan University.

Bley, J. E., (1983). Use of an assessment center for selecting administrators and diagnosing inservice needs. Doctoral dissertation, Northern Arizona University.

Breed, R. D. (1985). Concurrent validity of two administrator selection procedures. Doctoral Dissertation, University of Nebraska.

Britt, W. C. (1976). A comparative study of on-the-job ratings to assessment center ratings of school administrators. Doctoral dissertation, Georgia State University.

Crumpton, S. L. (1989). The importance of the national association of secondary school principals assessment center report to the appointment of public school principals in south carolina as perceived by superintendents and first-time principals. Doctoral dissertation, University of South Carolina.

Dickson, K. G. (1987). The oregon assessment center: Its influence on candidates' career opportunities and professional development. Doctoral dissertation, University of Oregon.

Douglas, J. R. & Johnson, M. C. Interrater reliability, interrater agreement, and bias in a nationally sanctioned assessment center. (Report No. EA 020 233). Wichita State University. (ERIC Document Reproduction Service No. ED 300879)

Draper, J. M. (1979). A comparative study of elementary school principals on-the-job ratings and their ratings from an assessment center designed to identify developmental needs. Doctoral dissertation, Georgia State University.

Ehinger, J. M. (1986). The validation of selected assessment center activities for incumbent school administrators who participated in the university of tulsa education management development program. Doctoral dissertation, University of Tulsa.

Engel, T. E. (1989). An analysis of the content and methods of instruction at michigan institutions that prepare principals. Doctoral dissertation, Western Michigan University.

Farmer, N. T. (1985). An analysis of the effects of participation in the national association of secondary school principals' assessment center in howard county, maryland. Doctoral dissertation, University of Maryland.

Ford, M. J. (1987). An attitude survey of NASSP principal assessment center candidates (Doctoral dissertation, Indiana University of Pennsylvania, 1987).

Friedman, M. J. (1980). Differences in assessment center performance as a function of the race and sex of ratees and the race of assessors. Doctoral dissertation, The University of Tennessee.

Heater, J. D. (1989). Teacher morale as a criterion of the predictive validity of the national association of secondary school principals assessment center. Doctoral dissertation, University of Nebraska.

Holman, David M. (1986). The relationship of participants performance in the nebraska assessment center to participant conceptual level/cognitive complexity. Doctoral dissertation, University of Nebraska.

Jacobson, B. J. (1986). Predictors of success for school administrators. Doctoral dissertation, University of Nebraska.

Light, Richard J. & Pillemer, David B. (1984). Summing up, the science of reviewing research. Harvard University Press, Cambridge MA.

Micek, T. J. (1983). Teacher perceptions of performance on the NASSP assessment center behavior dimensions as it relates to the conceptual complexity of principals. Doctoral dissertation, University of Nebraska.

Mullen, B. (1989). Advanced basic meta-analysis. New Jersey: Lawrence Berlaum Associates.

Ogawa, R. T., & Oxaal, I. (1986, April). A reexamination of nassp assessment center ratings: research notes. Paper presented at the Annual Meetings of the American Educational Research Association, San Francisco.

Pokorny, Nancy G. (1985). Techniques used in the selection of public school principals in the united states. Doctoral dissertation, University of Colorado.

Rosenthal, Robert (1989). Meta-analytic procedures for social research. Applied Social Research Method Series. Sage Publications, Newbury Park, CA (vol. 6).

Schrup, M. G. (1989). Role of reflection on administrative job performance following national association of secondary school principals' assessor training and assessment center service. Doctoral dissertation, Indiana University.

Schmitt N., Noe, R., Meritt, R., Fitzgerald, M., & Jorgensen, C. (1982). Criterion-related validity of the nassp assessment center. Unpublished report, Michigan State University, Department of Psychology, East Lansing, MI 48824).

Schmitt N., & Cohen, S. A. (1990). Criterion-related validity of the nassp assessment center. Unpublished report, Michigan State University, Department of Psychology, East Lansing, MI 48824).

Scholl, S. S. (1980). Criteria for selection of ohio middle school principals with particular attention to the leadership role in curriculum development. Doctoral dissertation, Miami University, Oxford, OH.

Storlien, H. T. (1983). Recruitment and selection of secondary principals in alberta: an analysis of procedures, criteria, and identified problems. Doctoral dissertation, University of Montana.

Thornton, G.C. III. and Byham, W.C. (1982). Assessment centers and managerial performance. New York: Academic Press.

Travetto, S. K. (1990). An analysis of school system and participant use of assessment center results for professional growth and development. Doctoral dissertation, University of Maryland.

Tryjankowski, E. M. (1990). Reliability and validity of the assessment and development center of new jersey. Trenton: New Jersey School Boards Association.

Van Newkirk, C. J. (1984). A study of the assessment center as a process for identifying prospective school principals and assistant principals. Doctoral dissertation, College of William and Mary.

Voit, L. B. (1989). Perceptions and evaluations of university principal preparation programs by michigan public school principals. Doctoral dissertation, Western Michigan University.

Walden, D. L. (1985). The effects of the NASSP assessment center process on self-improvement of assessees. Doctoral dissertation, University of Colorado.

Wendel, F. C., Gappa, L. R., & Yusten, C. R. Use of the assessment center method in administrator preparation programs. Tempe: Arizona State University, The University Council for Educational Administration.

Witters-Churchill, L. J. (1988). University preparation of the school administrator: Evaluations by texas principals. Doctoral dissertation, Texas A & M University.

Wolf, Fredric M. (1987). Meta-analysis, quantitative methods for research synthesis. Sage University Paper series on Quantitative Applications in the Social Sciences, 07-059. Beverly Hills: Sage Pubns.

Appendi
Description o

Author(s)	Form of Study/Year	Study Type	Sample Size
Alvy	Dissertation 1983	Associative Comparative	70
Blanck	Dissertation 1989	Comparative Descriptive	394
Bley	Dissertation 1983	Comparative Descriptive	120
Breed	Dissertation 1985	Comparative	19
Britt	Dissertation 1976	Comparative	24
Crumpton	Dissertation 1989	Comparative Descriptive	243
Dickson	Dissertation 1987	Comparative Descriptive	201
Douglas/ Johnson	Report	Comparative	144
Draper	Dissertation 1979	Comparative	36

x 1
f Studies

Criterion Type
Administrative Process/ Responsibilities
Skill Dimensions/Overall Rating
Attitude Rating
Administrator Perceiver Interview
Administrative Competency Rating (ARCS)
Ranking of Selection Criteria
Overall Rating
Skill Dimensions Overall Rating
Superordinates and Subordinates Rating

Author(s)	Form of Study/Year	Study Type	Sample Size
Ehinger	Dissertation 1986	Comparative	183
Engel	Dissertation 1989	Comparative Descriptive	5
Farmer	Dissertation 1985	Comparative Descriptive	125
Ford	Dissertation 1987	Comparative Descriptive	249
Friedman	Dissertation 1980	Comparative	256
Heater	Dissertation 1989	Comparative	152
Holman	Dissertation 1986	Associative Comparative	61
Jacobson	Dissertation 1986	Comparative Predictive	43
Micek	Dissertation 1983	Comparative Predictive	455
Ogawa/ Oxaal	Report 1986	Comparative Descriptive	139
Pokorny	Dissertation 1985	Descriptive	141

Criterion Type
Skill Dimensions/Overall Rating
Superintendent Rating
Universities Programs
Participant Responses
Participant Attitude
Skill Dimensions/Overall Rating
Purdue Teacher Questionaire
Performance on Paragraph Complexity Test
Educational Administration Grades/ Ratings
Paragraph Completion Test
Administrative Behavior Dimensions
Skill Dimensions/Overall Rating
Participant Responses

Author(s)	Form of Study/Year	Study Type	Sample Size
Schmitt (1) et al	National NASSP Study 1981	Comparative Predictive	39/ 350
Schmitt (2) Cohen	National NASSP Study 1990	Comparative Predictive	2699/ 2910
Scholl	Dissertation 1980	Comparative Descriptive	119
Schrup	Dissertation 1989	Descriptive	16
Storlien	Dissertation 1983	Comparative Predictive	372
Travetto	Dissertation 1990	Comparison Descriptive	415
Tryjankowski	Dissertation 1990	Comparative Predictive	65
Van Newkirk	Dissertation 1984	Qualitative	
Voit	Dissertation 1989	Comparative Descriptive	166

Criterion Type
Skill Dimensions/Overall Rating Performance Rating
Skill Dimensions/Overall Rating Performance Rating
Participant Criterion Ranking
Interview and Rating Scale of Skill Dimensions
Selection Criteria
Use of Assessment Center Results
Skill Dimensions/Overall Rating
Historical Analysis
Ranking of Skill Development Levels

Author(s)	Form of Study/Year	Study Type	Sample Size
Walden	Dissertation 1985	Qualitative	12
Wendel/ Gappa/ Yusten	Research Report	Comparative	225
Witters-Churchill	Dissertation 1988	Comparative Descriptive	325

Criterion Type
Congruency Rating
Survey of Use of Assessment Center Methods
Ranking of Skill Development Levels

Appendix 2
Categorization of Study Types

Author	PRECURSORS Exclusively Descriptive	PROCESS DIMENSIONS Exclusively Descriptive/Inferential	CRITERIA Exclusively Descriptive/Inferential
Alvy			X
Blanck	X	X	X
Bley		X	X
Breed		X	X
Britt			X
Crumpton		X	X
Dickson	X	X	X
Douglas	X	X	X
Draper			X
Ehinger		X	X
Engel	X		X
Farmer		X	X
Ford	X	X	X
Friedman		X	
Heater			X

Author	PRECURSORS Exclusively Descriptive	PROCESS DIMENSIONS Exclusively Descriptive/Inferential	CRITERIA Exclusively Descriptive/Inferential
Holman			X
Jacobson		X	X
Micek			X
Ogawa		X	
Pokorny	X		X
Schmitt (1) & (2)	X	X	X
Scholl		X	X
Schrup		X	X
Storlien			X
Travetto	X		X
Tryjankowski	X	X	
Van Newkirk		X	
Voit	X	X	
Walden			X
Witten-Churchill	X	X	
Wendel	X		

Appendix 3
Variables In The Meta-Analytic Model

PRECURSORS

	DESCRIPTIVE ONLY
ADMINISTRATION/ PREPARATION PROGRAMS:	Engel, Voit, Wendel, Witters-Churchill
AGE:	Bley, Crumpton, Ford
DEGREE:	Douglas, Pokorny
GENDER:	Blanck, Bley, Crumpton, Dickson, Douglas, Ford, Friedman, Holman, Pokorny, Schmitt, Scholl, Travetto, Tryjankowski
POSITION TYPE:	Bley, Blanck, Dickson, Douglas, Farmer, Ford, Holman, Schmitt, Tryjankowski, Voit, Witters-Churchill
RACE:	Blanck, Crumpton, Dickson, Douglas, Friedman, Ford, Pokorny, Schmitt, Scholl, Travetto, Tryjankowski
REGION:	Alvy, Ford, Pokorny
SENIORITY:	Bley, Douglas, Pokorny
SIZE OF DISTRICT:	Crumpton, Pokorny, Travetto
YEARS OF EXPERIENCE:	Bley, Douglas, Ford, Friedman, Heater, Holman, Micek, Pokorny, Travetto, Walden

PROCESS

	EXCLUSIVELY DESCRIPTIVE
SESSMENT PROCESS:	Blanck
SESSMENT AINING/ACTIVITIES:	Ehinger, Schrup, Van Newkirk, Wendel
ERALL RATING:	
SSP SKILL MENSIONS:	Breed, Ford, Scholl Voit, Witten-Churchill

DESCRIPTIVE AND INFERENTIAL

Draper

Ford, Schmitt

Bley, Dickson, Ehinger,
Ehinger, Ford, Holman,
Jacobson, Ogawa, Schmitt

	CRITERIA EXCLUSIVELY DESCRIPTIVE
ALTERNATIVE ADMINISTRATIVE PROGRAMS:	Engel
ATTITUDE RATING:	Farmer, Schrup, Travetto, Walden
CONCEPTUAL RATING:	
LEGAL ASPECTS:	Van Newkirk
OVERALL RATING:	Farmer, Travetto
PARTICIPANT/ NON-PARTICIPANT:	Bley
PERFORMANCE RATING:	Alvy, Breed, Farmer, Schmitt, Van NewKirk, Walden
POSITION ASSIGNMENT:	Bley, Dickson, Douglas Storlein, Travetto, Tryjankowski
SELECTION CRITERIA:	Engel, Pokorny Scholl, Storlein
ALTERNATIVE SKILL DIMENSIONS:	Dickson, Engel, Farmer Scholl
TEACHER MORALE:	Heater

Variables in bold print indicate effect sizes combined

DESCRIPTIVE AND INFERENTIAL

Schrup

Bley, Ford, Dickson
Jacobson

Holman, Micek

Ford

Britt, Draper, Ehinger,
Jacobson, Schmitt

Draper, Holman, Jacobson

Crumpton

Blanck, Bley, Breed, Douglas,
Ehinger, Jacobson

in the meta-analysis.